

```
In [1]: 1 import requests
        2 from bs4 import BeautifulSoup
        3 import pandas as pd
```

```
In [2]: 1 url = 'https://www.themoviedb.org/movie?page='
        2 page_url = 'https://www.themoviedb.org'
```

```
In [3]: 1 content=requests.get(url,headers={'User-Agent':'Mozilla/5.0'}).text
```

```
In [4]: 1 ## movie_url of 1st page
        2 soup = BeautifulSoup(content,'lxml')
        3 movie_url_1st_page = []
        4 movie_name_lst = []
        5 data = soup.find_all('div',class_ = 'card style_1')
        6 for i in data:
        7     movie_code = i.a['href']
        8     movie_url_1st_page.append(page_url+movie_code)
        9     names = i.a['title']
       10     movie_name_lst.append(names)
```

```
In [5]: 1 director_lst = []
2 gen_lst = []
3 run_time = []
4 release_lst = []
5 raiting_lst = []
6
7 for link in movie_url_1st_page:
8     content=requests.get(link,headers={'User-Agent':'Mozilla/5.0'}).text
9
10    soup = BeautifulSoup(content,'lxml')
11    data = soup.find_all('div',class_ = 'header_poster_wrapper true')
12
13    raiting = soup.find('div',class_ = 'user_score_chart')['data-percent']
14    raiting_lst.append(raiting)
15
16
17    release_date = soup.find('span',class_ = 'release').text.split()[0]
18    release_lst.append(release_date)
19
20    director = soup.find('li',class_ = 'profile').a.text
21    director_lst.append(director)
22
23    val = str(soup.find('span',class_ = 'genres').text)
24    genres = val.replace('\n','')
25    gen_lst.append(genres)
26
27
28    runtime = str(soup.find('span',class_ = 'runtime').text)
29    time = runtime.replace('\n','')
30    run_time.append(time)
31
32    movie_data_dic = {
33        'Movie Name': movie_name_lst,
34        'Raiting' : raiting_lst,
35        'Release Date' : release_lst,
36        'Run Time': run_time,
37        'Genres': gen_lst,
38        'Director': director_lst,
39        'Movie_link': movie_url_1st_page
40    }
41 }
```

```
42  
43  
44  
45  
46  
47
```

```
In [6]: 1 Df = pd.DataFrame(movie_data_dic)
```

In [7]:

1 Df

Out[7]:

	Movie Name	Raiting	Release Date	Run Time	Genres	Director	Movie_link
0	Doctor Strange in the Multiverse of Madness	75.0	05/06/2022	2h 6m	Fantasy, Action, Adventure	Steve Ditko	https://www.themoviedb.org/movie/453395
1	Fantastic Beasts: The Secrets of Dumbledore	68.0	04/15/2022	2h 22m	Fantasy, Adventure, Action	David Yates	https://www.themoviedb.org/movie/338953
2	Dog	74.0	02/18/2022	1h 42m	Drama, Comedy	Reid Carolin	https://www.themoviedb.org/movie/626735
3	Sonic the Hedgehog 2	77.0	04/08/2022	2h 2m	Action, Adventure, Family, Comedy	Josh Miller	https://www.themoviedb.org/movie/675353
4	Morbius	65.0	04/01/2022	1h 45m	Action, Science Fiction, Fantasy	Daniel Espinosa	https://www.themoviedb.org/movie/526896
5	The Lost City	68.0	03/25/2022	1h 52m	Action, Adventure, Comedy	Adam Nee	https://www.themoviedb.org/movie/752623
6	Spider-Man: No Way Home	81.0	12/17/2021	2h 28m	Action, Adventure, Science Fiction	Steve Ditko	https://www.themoviedb.org/movie/634649
7	Memory	73.0	04/29/2022	1h 54m	Action, Thriller, Crime	Martin Campbell	https://www.themoviedb.org/movie/818397
8	Collision	60.0	06/16/2022	1h 39m	Thriller, Crime, Drama	Fabien Martorell	https://www.themoviedb.org/movie/975714
9	Centauro	66.0	06/15/2022	1h 29m	Action, Crime, Thriller	Daniel Calparsoro	https://www.themoviedb.org/movie/852592
10	Spiderhead	57.0	06/17/2022	1h 46m	Science Fiction, Thriller	Joseph Kosinski	https://www.themoviedb.org/movie/615469
11	Jurassic World Dominion	67.0	06/10/2022	2h 27m	Action, Adventure, Science Fiction	Colin Trevorrow	https://www.themoviedb.org/movie/507086
12	The Black Phone	72.0	06/24/2022	1h 43m	Horror, Thriller	Scott Derrickson	https://www.themoviedb.org/movie/756999
13	Shark Bait	70.0	05/13/2022	1h 27m	Horror, Thriller, Action	James Nunn	https://www.themoviedb.org/movie/960258
14	Panama	59.0	03/17/2022	1h 40m	Action, Thriller	Mark Neveldine	https://www.themoviedb.org/movie/628878

	Movie Name	Raiting	Release Date	Run Time	Genres	Director	Movie_link
15	The Northman	73.0	04/22/2022	2h 17m	Action, Adventure, Fantasy	Robert Eggers	https://www.themoviedb.org/movie/639933
16	Turning Red	75.0	03/10/2022	1h 40m	Animation, Family, Comedy, Fantasy	Domee Shi	https://www.themoviedb.org/movie/508947
17	Uncharted	71.0	02/18/2022	1h 56m	Action, Adventure	Ruben Fleischer	https://www.themoviedb.org/movie/335787
18	The Desperate Hour	61.0	09/12/2021	1h 24m	Thriller	Phillip Noyce	https://www.themoviedb.org/movie/764835
19	Hustle	79.0	06/03/2022	1h 58m	Drama, Comedy	Jeremiah Zagar	https://www.themoviedb.org/movie/705861

In []:

1

In [8]:

```

1 url_lst = []
2 for u in range(0,501):
3     url_lst.append(url+str(u))

```

In []:

1

In [9]:

```

1 movie_url_all_pages = []
2 movie_name_lst = []
3
4 for link in url_lst:
5     content=requests.get(link,headers={'User-Agent':'Mozilla/5.0'}).text
6     soup = BeautifulSoup(content,'lxml')
7     data = soup.find_all('div',class_ = 'card style_1')
8
9     for i in data:
10         movie_code = i.a['href']
11         movie_url_all_pages.append(page_url+movie_code)
12         names = i.a['title']
13         movie_name_lst.append(names)
14

```

```
In [10]: 1 len(movie_name_lst)
```

```
Out[10]: 10000
```

In [33]:

```
1 director_lst = []
2 gen_lst = []
3 run_time = []
4 release_lst = []
5 raiting_lst = []
6
7 for link in movie_url_all_pages:
8     content=requests.get(link,headers={'User-Agent':'Mozilla/5.0'}).text
9
10
11     soup = BeautifulSoup(content,'lxml')
12     data = soup.find_all('div',class_ = 'header_poster_wrapper true')
13
14     raiting = soup.find('div',class_ = 'user_score_chart')['data-percent']
15     raiting_lst.append(raiting)
16
17
18     release_date = soup.find('span',class_ = 'release').text.split()[0]
19     release_lst.append(release_date)
20
21     director = soup.find('li',class_ = 'profile')
22     if director is not None:
23         director=(director.p.text)
24     director_lst.append(director)
25
26     val = str(soup.find('span',class_ = 'genres').text)
27     genres = val.replace('\n','')
28     gen_lst.append(genres)
29
30     runtime = soup.find('span',class_='runtime')
31     if runtime is not None:
32         runtime=(runtime.text.strip())
33         #run_time.append(runtime)
34
35     run_time.append(runtime)
36
37     movie_data_dic = {
38         'Movie Name': movie_name_lst,
39         'Raiting' : raiting_lst,
40         'Release Date' : release_lst,
41         'Run Time': run_time,
```

```
42         'Genres': gen_lst,  
43         'Director': director_lst,  
44         'Movie_link': movie_url_all_pages  
45     }  
46 }
```

```
In [34]: 1 len(run_time)
```

```
Out[34]: 10000
```

```
In [35]: 1 len(director_lst)
```

```
Out[35]: 10000
```

```
In [36]: 1 len(movie_url_all_pages)
```

```
Out[36]: 10000
```

```
In [37]: 1 len(release_lst)
```

```
Out[37]: 10000
```



```
In [38]: 1 df = pd.DataFrame(movie_data_dic)
          2 df
```

Out[38]:

	Movie Name	Raiting	Release Date	Run Time	Genres	Director	Movie_link
0	Doctor Strange in the Multiverse of Madness	75.0	05/06/2022	2h 6m	Fantasy, Action, Adventure	Steve Ditko	https://www.themoviedb.org/movie/453395
1	Fantastic Beasts: The Secrets of Dumbledore	68.0	04/15/2022	2h 22m	Fantasy, Adventure, Action	David Yates	https://www.themoviedb.org/movie/338953
2	Dog	74.0	02/18/2022	1h 42m	Drama, Comedy	Reid Carolin	https://www.themoviedb.org/movie/626735
3	Sonic the Hedgehog 2	77.0	04/08/2022	2h 2m	Action, Adventure, Family, Comedy	Josh Miller	https://www.themoviedb.org/movie/675353
4	Morbius	65.0	04/01/2022	1h 45m	Action, Science Fiction, Fantasy	Daniel Espinosa	https://www.themoviedb.org/movie/526896
...
9995	The Scary House	57.0	10/30/2020	1h 40m	Family, Fantasy, Horror	Daniel Prochaska	https://www.themoviedb.org/movie/735595
9996	Maria Full of Grace	73.0	04/01/2004	1h 41m	Drama, Thriller, Crime	Joshua Marston	https://www.themoviedb.org/movie/436
9997	The Art of Getting By	66.0	06/17/2011	1h 23m	Drama, Romance	Gavin Wiesen	https://www.themoviedb.org/movie/64678
9998	I'm a Girl, I'm a Princess	60.0	10/28/2021	2h	Drama	Federico Palazzo	https://www.themoviedb.org/movie/812491
9999	Lake Eerie	38.0	01/15/2016	1h 44m	Thriller, Horror, Science Fiction	Chris Majors	https://www.themoviedb.org/movie/316268

10000 rows × 7 columns

```
In [41]: 1 df.to_csv("Movie_Data.csv")
```

In []:

1
2

In []:

1

In []:

1

In []:

1